Centre for Policy Modelling, MMUBS, Manchester
Stockholm Environment Institute, Oxford Office
Universität Kassel
Politechnika Wroclawska
IIASA, Vienna
Macaulay Institute, Aberdeen
Uniwersytet Wroclawski

Case Study Validation Protocol

Deliverable No. 10 of Project 012816: CAVES – Complexity, Agents, Volatility, Evidence and Scale

Duration: 2005-2008

Funded under the EU 6FP NEST programme.

Table of Contents

1 Introduction

The CAVES Project aims to develop models and modelling procedures that will assist local stakeholders in the formulation of land-use and related social policy under changing conditions. Given the use of these models in public policy, it is important for the modellers to establish the degree of confidence that can be placed in their findings. The case studies focus on three regions, the Grampian in Scotland, the Odra River Valley in Poland, and the Limpopo Province in South Africa. Land use in each of the case study areas is subject to a form of change. The three regional research teams have identified and established links with stakeholders in each of their respective regions and collected qualitative and quantitative information that has, in principle at least, informed the models developed by the respective modelling teams.

In practice the interaction between modellers and field-researchers has been defined by the regional circumstances and has included both integrated modelling and case study teams and separate modelling and case study teams.

- *Integrated Modelling and Case Study Teams*: In this type of modelling, modellers and fieldworkers have formed a cohesive team with strong interaction. Such interaction has emerged where the modelling team has been based in the case study region, such as was the case for the Macaulay Institute at Aberdeen, which has a history of land-use research in the region.

  This approach has the advantage of allowing fieldworkers to scrutinise and test the appropriateness of model assumptions, and modellers to discuss assumptions with fieldworkers on a day-to-day basis, allowing greater involvement with model development.

- *Separate Modelling and Case Study Teams:* This situation is typically characterised by distinct modelling and fieldwork teams often located in different places. In the Odra River Valley case study, for example, the case study team comprises of ecologists from a nearby university in Wroclaw, Poland, while the modelling team is from Kassel in Germany. The case study team is responsible for providing information to the modelling team. A similar situation exists in the South African case study. The research team is overseen by the Oxford, UK office of the Stockholm Environment Institute.

In both instances the reconciling and prioritisation of objectives is essential. Social simulation modellers tend to be interested in the complexity that characterises the social reality. Agent-based modellers often focus on mapping the cognitive processes of social entities and how these are influenced by endogenous and exogenous factors.

Where modellers have limited exposure to the case study region, reconciling these differences in emphasis and focus can be more difficult. The case study teams have approached the challenge of validation differently, typically having to rely on the resources and data available and operating within strict time and budget constraints. The approaches adopted draw, in many ways, on the theoretical and conceptual ideas discussed in the complementary CAVES validation concept paper. This paper describes the three case studies and where possible highlights the use of formal modelling and validation approaches.

Each case study is described and discussed in six common themes:

i. Objective of the case study
ii. Brief description of the domain
iii. Stakeholder participation activities and emergent findings
iv. Design of agents in case study model(s)
v. Validation data
vi. Development of stakeholder/data validation exercises and outcomes

2 Scotland: Grampian Hill Farms

2.1 Objective of the case study

The focus of the Grampian case-study is on land use change patterns in the Upper Deeside region of North East Scotland between the mid 1980s and 2030. For the period up to the present, both qualitative data from interviews, and quantitative data from existing datasets, are being used to build a case-study specific model within a refined version of the pre-existing FEARLUS modelling framework. Further quantitative information, and further interviews, will be used in an attempt to validate the model. Specific foci are on the drivers and processes of land use change, and the particular role of social networks in these processes. The Grampian case study was chosen to represent an area of relative stability in land use patterns in recent decades: despite apparent 'shocks' to the system, such as disease outbreak, subsidy restructuring and exit from the European Exchange Rate Mechanism. The Grampian region provides contrast to the Polish and South African sites, where changes have been more marked. The purpose of the Grampian case study is to draw on field research to inform and develop the existing FEARLUS (Framework for the Evaluation and Assessment of Regional Land Use) modelling framework, resulting in the generation of scenarios of possible future land use patterns. These scenarios are intended for use in discussions of policy formation.

2.2 Brief description of the domain

During the past twenty years a series of slow and subtle changes, including increased farm scale, increased specialization, more intense stocking, decreased use of inputs, and increase in environmental legislation, have taken place on farms in the study site. These changes are perceived to be in response to progressive market liberalisation, and the 'de-coupling' of subsidies from production through the Single Farm Payment, labour availability and increasing mechanisation. Estate managers report actively encouraging tenants to increase the scale of their operations and undertake environmental projects; this reflects a general impression that farm size must increase for operations to remain viable. Actual land uses remain fairly stable, (primarily beef, sheep and barley), reflecting a perceived lack of alternatives. Respondents expressed concern about the ongoing viability of farming in the area, as subsidy reductions occur over time. There is general belief that farms in 'good' farming regions (where there is considerable arable land) will continue to get larger, in order to benefit from economies of scale. In poorer regions, such as highland areas with little arable land, farmers expect agriculture to become more extensive (decreased use of inputs, slower maturing of livestock), and land abandonment to occur in remote regions.

2.3 Stakeholder participation activities and emergent findings

The Grampian case study seeks to include a wide range of land users. To this end, interviews were conducted with 24 farmers (29 including spouses or partners also participating), 6 identified successors (to farmers in the study), 5 estate managers, and 9 key informants in the Upper Deeside region. Agricultural land holding in Upper Deeside is mixed: towards the east, most farmers own their land; further west, the land is typically owned by estates, and held by farmers on a tenanted basis. Some farmers in the study held a mix of owned and tenanted land. Farm holdings in the study range from 30 to 5500 acres in size, although most are between 300 and 700 acres. Arable land typically comprises less than a third of most farms, with the bulk of land being utilised as improved pasture and grazing. All of the farmers produce beef or sheep, and livestock feed (silage and/or barley) with some also producing malting barley, potatoes, turnips and oil seed rape. Although most farmers in the study were reliant on their farm for their livelihood, they also typically had off-farm income, most commonly machinery contracting to other farms and spousal employment, but also including tourist activities. Most farm operations also participate in environmental programs as a source of revenue generation.

The interviews took place in four stages over the course of the project: pilot study, primary field research, validation interviews and final interviews, covering a period of two years.

The iterative approach to field research allows considerable feedback between interview data collection and the modelling process. As the interviews are qualitative in nature, there is a great deal of flexibility in the topics addressed, allowing the research to evolve according to new findings and modelling issues. It is not intended that research findings be formally representative of the study site; however, the research is structured to include a broad range of farm and estate types.

In addition to information about farmer decision-making, the interviews conducted so far have enabled us to identify a number of key trends relevant to our modelling approach over the period to be modelled, notably tendencies for farms to increase in size, reduce the number of commodities produced (for example, pigs, potatoes and dairy cattle have largely disappeared from the study area), and shift toward less labour intensive land uses. They have also enabled us to identify the major shocks and stresses over the period form the mid-1980s to the present:

- During the 1980s there was a rapid fall in land and output prices, with banks foreclosing on a large number of loans to farmers around 1985.
- 1992 saw the UK leave the ERM, leading to devaluation of the £ sterling. In the same year, the McSharry reforms to the CAP occurred, and IACS arable aid payments were introduced. These developments caused a sharp land price rise to 1995, which continued more gradually as cereal prices dropped.
- In 1996, news that BSE could be transmitted to humans destroyed the export markets for beef and breeding stock. Note that there were few cases in Scotland: the shock felt by Scottish farmers did not stem directly from the cattle disease (i.e. their cattle did not die in large numbers), but from resulting price and regulatory changes.
- The 2001 foot and mouth epidemic did not reach the area, but control measures affected tourism and to a lesser extent farming.
- The approach of the CAP switch to the Single Farm Payment caused great uncertainty. Effects of the change itself have so far been less marked than expected, as they coincided with a rise in product prices.
- Shortage of labour is a considerable constraint on the choice of land management strategies available to Grampian farmers.

With the exception of labour shortages, these shocks and stresses will have operated almost entirely through changes in prices, and so can be incorporated in FEARLUS model runs to the extent that reliable price data is available. We are still considering how labour shortage could best be incorporated in the model.

2.4 Design of agents in case study model(s)

The environment created by the FEARLUS modelling framework is populated by Land Managers (there is also a Government agent). Each Land Manager owns at least one Land Parcel. The set of Land Parcels managed by one particular Land Manager is called an Estate. At any particular time, each Land Parcel is managed by one (and only one) Land Manager, who has allocated one (and only one) certain Land Use in the Land Parcel, using a particular decision-making algorithm. Land Managers undertake various activities every time-step (which is intended to represent one Year):

a) Land Managers decide a Land Use for each of the Land Parcels in their Estate (“Land Use decision making”).

b) At the end of the Year, Land Managers accumulate their annual income, which affects their Wealth (“Income generation”).

c) Some classes of Land Managers may approve or disapprove of other Land

Managers in their social neighbourhood ("Social approval" and "social disapproval"). This social approval or disapproval can affect some Land Managers' decisions on what Land Use to select for certain Land Parcels.

d) "Learning". Whether or not this happens depends on the decision making algorithm.

e) Depending on their Wealth at the end of the Year, Land Managers may decide to buy Land Parcels ("Land Parcel transfer").

Land Managers are grouped into Subpopulations, which potentially make decisions in different ways. Even when two Land Managers belong to the same Subpopulation (i.e. have been created using the same parameter file), they can be substantially different. This is so because the Subpopulation files may specify probability distributions for parameter values, rather than exact values.

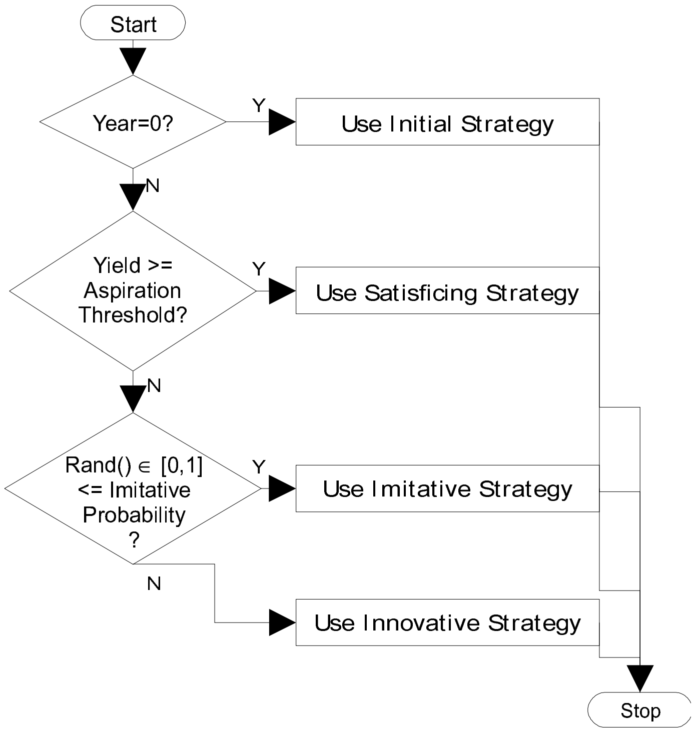

Figure 1:Flowchart of underlying structure of the Land Use Decision Algorithm employed by Land Managers belonging to the SubPopulation class.

Most Subpopulation classes implement a Land Use decision making algorithm that can be summarised in two steps: first, a decision to reconsider Land Use; second, if reconsideration is undertaken, the selection of a Land Use (which may turn out to be the same as the former one). The decision to reconsider Land Use may involve an assessment of: (i) whether the Land Manager's Wealth Aspirations are met, (ii) whether the Land Manager's Social Approval Aspirations are met, (iii) the period of time over which these Aspirations have/have not been met. The use of Aspirations mirrors Simon's (1955) theory of decision-making based on 'satisficing': beyond a certain point, people will not make changes because what they have is 'good enough'—the risk of making changes is too great relative to the perceived potential gains. Thus, Land Managers only change Land Use when Aspirations are not met (for a sufficient period of time). The choice of new Land Use may be based on one of four methods: (i) An 'innovative' or 'experimental' method, which assumes the Land Manager knows of all possible Land Uses that might be applied, and has criteria for choosing among them heuristically; (ii) An 'imitative' method, using similar criteria but assuming only knowledge of Land Uses occurring in the spatial neighbourhood of the Land Manager's Estate; (iii) Case-Based Reasoning (Aamodt & Plaza, 1994), a method derived from cognitive psychology allowing decisions to be based on prior experience of a Land Use; (iv) Advice—a method that can be integrated into the CBR approach to ask neighbouring Land Managers about their experiences with a Land Use when the Land Manager has none themselves. Probability distributions are used within a Subpopulation to give Land Managers an individual propensity to base their decision to change on any of the three factors outlined above, and to use any of the four outlined methods to make the change. More detail on Land Manager decision-making can be found in the FEARLUS 1.0 User Guide (Polhill and Izquierdo, 2007).

The Social Approval model assumes that Land Managers have a set of Triggers influencing whether or not they may Approve or Disapprove of each other. This is based on evidence from the literature that different subtypes of farmer have various set notions about management practices that do or do not indicate a 'good farmer' (Burton, 2004). Land Managers' Triggers are thus effectively rules indicating whether to approve or disapprove of a neighbour using specific Land Uses.

Land Parcel transfer is discussed in detail in Polhill, Parker and Gotts (2005; 2008). Land Managers with negative accumulated Wealth are assumed to be bankrupt, and put all their Parcels up for Sale. These Parcels are then auctioned to neighbouring Land Managers, and a potential in-migrant Land Manager (the Subpopulation of which is chosen using probabilities that may vary over time). Land Managers have heuristic Strategies for deciding which Land Parcels to bid for, and how much to offer.

2.5 Validation data

The FEARLUS modelling framework was in development for several years prior to the CAVES project, with the structure informed largely by literature review. What the Grampian field research has brought to the FEARLUS model is data specifically derived for use in the model, i.e. a form of docking or TAPAS (Take A Pre-existing model and

Add Something). Due to the focus of CAVES on complexity and social networks, these are the areas in which the model has most advanced.

2.5.1 Validation using qualitative data

There are no standard methods for utilising qualitative research in agent-based modelling. More generally, standard qualitative research validation techniques include the utilisation of multiple data sources (for example interviewees, key informants, data from other research projects and published literature), and feedback on research results by research respondents. Utilising best practice is an academic standard of data validation, and is also part of the validation protocol.

For the purposes of validating qualitative inputs to the FEARLUS model, a questionnaire was derived from model components and assumptions. These include the comprehensiveness and relative importance of factors in land use change, principles of land use change, and the decision-making process. As agent-based models are not intended to be predictive, outputs of the model were not given to respondents for validation. The focus was instead on model inputs and processes. Consistent with the TAPAS approach, emphasis in validation is on demonstrating improvements to the model resulting from CAVES field research. Similarly, the particular focus on qualitative aspects of model processes (such as decision-making and social approval) reflects the primary input of the field research into the model. However, it was possible to utilise the opportunity of the validation exercise to validate the range of factors included in the calculating farm profitability, and the classification used for agricultural land.

2.5.2 Validation using quantitative data

For quantitative data, validation is possible through cross referencing with existing datasets and prior research. The intention is to use available quantitative information on the changes in land use in the study area (and information on farm size change so far as this can be obtained) for input calibration and macro-validation of the model as applied to the past. Since the actual course of events must be considered nondeterministic, no single run of the model could be expected to reproduce the actual course of events precisely; besides which, the quantitative data available on both inputs (climatic, land suitability and price data, and data on the effects of climate and land suitability on yields) and outputs (land use and farm size data) is limited, and in some case affected by confidentiality restrictions. Therefore we will be assessing whether the model is able to reproduce the direction and magnitudes of the trends found in the data concerning land use and farm size, given the best available data relevant to model inputs.

The general approach being taken is as follows (these steps are overlapping rather than sequential; all but the last two are underway):

- Select those aspects of the world that can be represented in some way by inputs or outputs of the model. Some of these aspects (e.g. farmer decision-making procedures, climatic and economic conditions, available land uses) are inputs to the model; others (land use distribution and farm size) are outputs.

- For each of these aspects, determine what data are available for the period from the mid-1980s to the present. Farmer decision-making is discussed below under the heading of stakeholder/data validation exercises; other aspects are discussed here.
- Where there are data relevant to input parameters, determine how it can best be encoded in those parameters.
- Where there are no data good enough to be worth using for a particular input parameter, select a range of plausible combinations of parameter values with which to run the model.
- Select an area similar to, but not too close to, the case study area. Use this to calibrate those aspects of the model (primarily concerning land manager decision-making) which cannot be estimated from the data sources described below. (Ideally, data used for calibration and validation should be independent; selecting a functionally similar but non-adjacent area for calibration is the closest we can get to this ideal.) This calibration exercise should narrow the range of plausible combinations of parameter values, since any combinations that produce clear mismatches with the real-world output data can be dropped.
- For each remaining plausible combination, run the model on the case study area data, as many times as is feasible. For each important quantitative variable adequately documented in the data sources available (e.g., change over a specific span of time in the number of farms in the study area, or in the proportion of land devoted to different land uses, we will be testing whether the variable's real-world value is within the range of values produced by ensembles of model runs: any such real-world value more extreme than 95% of an ensemble of runs would suggest a possible flaw in the model, and merit detailed investigation.

The remainder of this subsection lists the main kinds of model input and output to be considered, and the main sources of quantitative data. While far from ideal, these do appear sufficient to undertake quantitative validation work.

Model inputs and outputs

- Inputs needed. Input parameters specify:
    - Initial land uses.
    - Initial farm boundaries, size distribution.
    - Land capability.
    - Climatic conditions.
    - Expected yields under the relevant combinations of land quality and climatic condition.
    - Economic conditions, primarily prices for farm products and the inputs necessary to produce them.
    - Land use selection strategies.
    - Land purchase strategies.

    - Land managers' social interactions (approval or disapproval of neighbours, and the importance a land manager attaches to their neighbours' good opinion, relative to that assigned to financial considerations).
    - Off-farm income.
    - Economies of scale.
    - Government policies.
- Outputs that could be compared with empirical evidence (in all cases, both time series and spatial distributions need to be considered).
    - Land uses
    - Farm boundaries, farm size distribution
    - Land prices

Data on land uses and farm sizes

The most readily available and processed data on land use and farm size come from the EDINA transformation of the annual Agricultural Census data (http://edina.ac.uk/agcensus/). Land uses are given at various levels of specificity (e.g. numbers are given both for "total cattle", and categories as specific as "bulls for service"). For use with the current version of FEARLUS, broad categories are considered to be most suitable. Land use data is georeferenced using 10km, 5km and 2km squares, although created from parish-level data: the 2km square data will be used in quantitative validation. Data about farm size is actually expressed as the number of farms – and of farms of different types – within each grid square. More detailed data on farm size will be difficult to obtain: land sales are recorded in public document repositories, but not in a readily accessible way; and tenancy agreements do not appear to be so recorded.

Doubts have been raised about the quality of the EDINA data, in particular about the effect of the parish-to-grid transformation, but since much of our other data (see below) is also in gridded form, it is a considerable advantage to have land use and farm size data in this format.

Additional information concerning farm size (specifically, changes in farm sizes over time) may be drawn from Scottish Agricultural College's annual Farm Management Handbook (Beaton, Catto and Kerr 2007), which makes use of Farm Accounts Scheme sample data.

Land capability data

The EDINA version of Agricultural Census information also provides (at a 1km^2 scale) a division of land into a small number of categories which it describes as "land use", but which are actually a combination of information about land capability and availability for agricultural and other uses: the categories used for Scotland being "urban", "water", "upland", "woodland", "agricultural land (Scotland)" and "restricted agricultural land" (golf courses and similar areas). This is judged to be sufficient for initial work with FEARLUS.

Climate data

The UK Met Office provides free access to sufficient weather station data for FEARLUS purposes, back to 1961. Since FEARLUS works on an annual timescale, the annualised data are the most obviously relevant. This includes, at a $5km^2$ scale:

1) Annual extreme temperature range (Highest daily maximum - lowest daily minimum) (°C)
2) Heating degree days - S (15.5 - daily mean temperature), ignoring negatives, over the winter months.
3) Growing degree days - S (daily mean temperature - 4), ignoring negatives, over the summer months.
4) Growing season length - Bounded by daily mean temperature being >5 °C for >five consecutive days and <5 °C for >five consecutive days.
5) Summer heat wave duration - Total days with maximum temperature >3 °C above the 1961-90 normal for >five consecutive days (May to October)
6) Winter heat wave duration - Total days with maximum temperature >3 °C above the 1961-90 normal for >five consecutive days (November to April)
7) Summer cold wave duration - Total days with minimum temperature >3 °C below the 1961-90 normal for >five consecutive days (May to October)
8) Winter cold wave duration - Total days with minimum temperature >3 °C below the 1961-90 normal for >five consecutive days (November to April)
9) Maximum number of consecutive dry days (days with less than or equal to 2 mm of rain) in a year
10) Greatest five-day precipitation total in a year (mm)
11) Rainfall intensity (annual rainfall total on rain days (days with >1 mm of rain) / number of rain days)

Since the annual data are drawn from a limited set of locations, interpolation will be necessary if the area of interest is not sufficiently near to a station. Which of these types of data are most relevant, and whether they provide sufficient information between them, depends on the particular land uses being modelled. So far as the upper Deeside study area is concerned, the most relevant are "growing season length", "summer heat wave duration", and "maximum number of consecutive dry days". All these affect the growth of grass: a long, warm, wet summer will produce the best growth, but a hot, dry summer will be particularly poor.

Data on yields

Data on yields, given land use, climatic conditions and land quality at the levels of detail appropriate to FEARLUS, have not been easy to find. The best general source appears to be Scottish Agricultural College's annual Farm Management Handbook. This series should be particularly useful with regard to changes in expected yield over time. Yield figures are always accompanied by a specification of the physical conditions and management practices assumed, and often in the form of high, medium and low estimates; this should make it feasible to adjust figures to take into account local conditions.

Data on economic conditions

The most readily useable source of information on both input and output prices are the time series produced by the UK's Department for Food and Rural Affairs (DEFRA). These series also include information on land prices (which could be used for comparison with those produced by the model, and potentially as a guide to incorporating exogenous influences on land prices); and of labour and fixed capital: factors not currently represented in FEARLUS, but potentially useful in interpreting simulation results. The Farm Management Handbook also contains price data on inputs, outputs and land, and these can be used as a check on the DEFRA prices.

2.6 Development of stakeholder/data validation exercises and outcomes

2.6.1 Validation protocol

The Grampian region validation protocol focuses on validation of the modelling framework components: comprehensiveness and relative importance of factors in land use change, principles of land use change, and the decision-making process. The focus is on model inputs and processes, rather than outputs. Heuristically, the FEARLUS models generated through the CAVES project follow the TAPAS approach to model design. Emphasis in validation is therefore on demonstrating improvements to the model resulting from CAVES field research. Similarly, the particular focus on social networks within CAVES is also reflected in the validation protocol.

In order to validate primary field research findings, questionnaires were completed with 8 respondents from the primary field research (5 farmers and 3 key informants). The farmers were selected to represent a range of tenures (tenanted and owner operated), ages (41-70), and approaches to farming (traditional vs. expanding). The key informants were also chosen to reflect different perspectives. These individuals were a farm business advisor, an environmental program advisor, and a representative of the National Farmers' Union. No factors (estate managers) or farming successors were drawn on for the validation, as the variety of their responses in the primary field research suggested that a single representative of either group would be unlikely to accurately represent, and thus validate, the perspective of the whole.

The questionnaire involved a variety of approaches, including open ended questions, ranking of identified factors and weighting of responses. A concurrent knowledge transfer project on the CAVES project also had a validation component, the results of which are also included in this report. In this project, agricultural industry members attending agricultural events (and therefore not formal study respondents) were asked to comment on the accuracy of statements made in an information note based on study findings.

2.6.2 Findings

The validation process generated broad support for the accuracy of the field research, while raising issues for further exploration. Respondents did not identify any major problems with the model inputs and process, and confirmed several aspects which were intended for addition to the model.

Validated model components:

- the decision rule that farmers do not change their current crop or type of stock when their aspiration threshold has been reached, even if there are higher prices in a different commodity.
- the general principle of innovation by a small number of innovative farmers, copied by other farms
- in ordering the factors which farmers take into consideration when changing commodity, the profitability of the new commodity is of primary importance.
- land is differentially desirable, on the basis of previous (and therefore anticipated) profitability.
- In historical runs, farmers will always bid on neighbouring land, if they have sufficient resources. This is less likely to be true in future-based runs.
- Farmer types: entrepreneurial, traditional, pluriactive, lifestyle/hobby/environmental

Important to add to the model:

- The principle that farmers will not always switch out of a commodity when they lose money on it, although restrictions to this principle on the basis length of time over which loss occurs and extent of loss have not been quantified.
- Off-farm income in the model, as a factor in agent's aspiration thresholds.
- Categorisation of agricultural land: 'arable', 'woodland', and 'improved grazing' are suitable categories. May want to change 'unimproved grazing' to 'rough or hill ground'.
- Farmer ability – impacts considerably on farm profitability.
- Including 'well established market for new commodity', 'suitability to current farm set-up' and 'opportunity to benefit from government grants' as factors in decision-making about new commodity up-take
- Including more specifics on proximity of land and expansion plans to decisions regarding land acquisition. (Expansion plans can be handled through farmer type; proximity is currently limited to the current geographic neighbourhood, rather than specifying immediate neighbours)
- Slow speed of farmer response to change events (e.g. one year delay before responding to the SFP)
- Climate change – will differentially impact by commodity and land use type.

Areas for further exploration:

- Fixed costs associated with commodities (not in model at present). Important, but there is not a linear relationship between number of commodities and fixed costs.
- Succession (not in the model at present) cannot be based solely on the initial success of the farm. Perhaps a complex issue best not addressed at this time.

- Organic farming (not an option in the model at present). Social approval may limit adoption, but land suitability and perceived economic benefits appear more important.
- Economies of scale can be utilised as a general principle, but it is also important to consider limitations imposed by other aspects of the farm business.
- No consistent identification of or response to major ‘shock’ events
- Social approval – difficult to accurately gauge.

3 South Africa: Rural Livelihoods in Sekhukhune District, Limpopo Province

3.1 Objective of the case study

The South African case study aims to represent rural livelihood security as related to water, health, employment and social networks in the Sekhukhune District Municipality, Limpopo Province in South Africa. Livelihoods in the region are impacted on by multiple stresses, including those caused by climatic variability, HIV/AIDS and limited employment opportunities. The hypothesis is that social networks within and between the village and municipal level affect the nature of adaptation to these stresses and hence how stresses are experienced.

The study intends to contribute to policy development by clearly articulating the links between policy and local and district social, economic and biophysical dynamics.

3.2 Brief description of the domain

The Sekhukhune District is an arid area and poor by South African standards. Water scarcity is a key stress at the district level and constrains daily village activity including field-based agriculture, home gardens, brick making projects and livestock husbandry. Social security grants and remittances from migrant labourers form the most common source of household income. Through the lens of climate, health and food security, this project has investigated how vulnerability is configured and what decisions people make regarding adaptation strategies in response to a multitude of stresses, which interact in complex and dynamic ways. The findings indicate that water scarcity and limited economic opportunities are two major constraints to development at both the village and district scale that undermines adaptive capacity (Figure 2). People relate both climate stress and food insecurity back to these two dominant stressors to a large extent. This highlights the need for integrated responses to support local adaptation that departs from hegemonic sectoral approaches. It also requires improved and increased communication between government and local communities to facilitate the integration of strategies being implemented at different scales and better align expectations. Municipal government needs to carefully assess the likely environmental, social and economic impacts of investing in different sectors, incorporating a view on climate change and prioritising water saving and wealth distributing options with the aim of minimizing future vulnerability.

Water for agriculture, industry and domestic use is provided by separate schemes, with the intention of guaranteeing water for domestic use. In reality all of these schemes are under pressure and many households do not have access to reticulated water. The recent rapid expansion of platinum and strategic minerals mining activities is an important contributor to the Sekhukhune economy but has exacerbated existing water shortages. Construction of the large De Hoop Dam and an expansion to the Flag Boshielo Dam are due to commence in 2007 but present additional environmental concerns.

Migrant labour to the mines and the metropolitan centres is a feature of Sekhukhune society and is associated with an increase in HIV/AIDS. Both migration and the spread of the HIV/AIDS undermine local social networks and contribute to high unemployment. Official 2003 statistics suggest 69.4% of the population is unemployed, but over 90% of the population is rural and in many rural villages the unemployment figure is much higher. In the context of unemployment there are suggestions from villagers that their inability, due to water shortages, to cultivate the same range of vegetables as previous generations and ensure local food security has undermined their health and contributed to poverty; an example of the dynamic social, environmental and economic nexus.

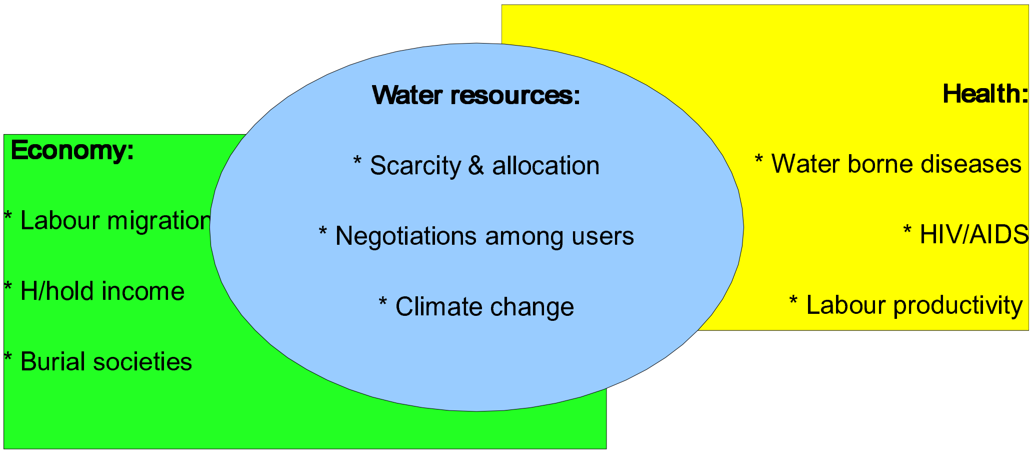

Figure 2: Schematic of areas of interest that emerged out of fieldwork to contribute to model development that links these to social dynamics

3.3 Stakeholder participation activities and emergent findings

Results from previous projects, including both qualitative and quantitative data, were used to inform and compile questions that were addressed in the 2006 pilot study. The first half of the field trip was mainly focused on household and focus group interviews.

Based on the initial results of this qualitative research, discrete choice analysis with stated preference data collection was conducted to compliment the qualitative data. Parallel to the ground level data collection, key person interviews were carried out with local authorities. The conclusion from the pilot study report, 'Adapting to Climate, Water and Health Stresses: insights from Sekhukhune, South Africa', highlights important issues for stakeholder engagement:

> "Focusing on water, climate, health and food security this project has investigated how vulnerability is configured and what decisions people make regarding adaptation strategies in response to a multitude of stresses, which interact in complex and dynamic ways. The findings indicate that water scarcity and limited economic opportunities are two major constraints to development at both the village and district scale that undermines adaptive capacity. People to a large extent relate both climate stress and food insecurity back to these two dominant stressors. This highlights the need for integrated responses to support local adaptation that departs from hegemonic sectoral approaches. It also requires improved and increased communication between government and local communities to facilitate the integration of strategies being implemented at different scales and better align expectations. Municipal government needs to carefully assess the likely environmental, social and economic impacts of investing in different sectors, incorporating a view on climate change and prioritizing water saving and wealth distributing options with the aim of minimizing future vulnerability."

The information gathered from this pilot study was used as starting point for the design of agent-based models. The purpose of the modelling was to analyse the role of social networks as support structures for communities dealing with environmental uncertainty as well as other types of change. Ga-Selala village was selected as the site where the study would be carried out. However, due to the diversity in village characteristics in the area, it is hard to know to what extent Ga-Selala can be regarded as a 'typical' village (see part v).

The early fieldwork reshaped the focus of the study from water and land management to the impact of mining, as it emerged that community and market gardens were not operational in the village as a result of lack of access to water, whereas mining was an increasing source of income and employment opportunities and therefore an important natural resource for the region.

The next sections provide discussion about follow-on work in 2007, which consisted of two further trips, that included fieldwork at the village level and with municipal representatives, and modelling work carried out in parallel - and integrated with - the case study work.

Later on in the project, as the modelling work of the MMU team progressed, the model-derived research issues and specific questions the field researcher could help to answer were brought into focus more. Therefore, for the second fieldwork trip (October 2007) the validation exercises were designed around model validation and presentation of model findings.

3.4 Design of agents in case study model(s)

Two models have been developed for the case study: a procedural model, which represents decision processes algorithmically, and a slightly more fine-grained declarative model, in which decision processes are governed by a set of rules. Both models are similar in that their agents represent individuals who are characterised by age, gender, marital status, health status etc. A group of (usually but not necessarily related) agents live together in a household, pooling and sharing all income and available resources. Each household has a head responsible for decisions that affect the household as a whole. Thus agents assume one of two different roles:

- Household member
- Household head

Agents belong to social networks, which both constrain their interactions with each other and evolve through these interactions. The models differentiate between family und kinship links, neighbours, friends, sexual partners and groups like churches, burial societies and savings clubs (stokvels).

Agents make decisions on the individual level and – if they function as a household head – on the household level. The latter concern the acquisition and distribution of food, borrowing food from relatives and/or neighbours if necessary, membership in burial societies, payment of school fees and the advertisement of piece jobs whenever the household is in need of and can afford hiring a worker. Decisions on the individual level regard applying for such piece jobs, migration to look for work outside the village, membership in stokvels, friends, sexual partners, marriage and building a house after marriage.

The declarative model makes use of so-called endorsements and individual tags in the agents' decision making. Endorsements can be thought of as labels used by an agent to describe certain aspects of other agents[1]. The model incorporates both positive labels like is-kin, is-neighbour, same-church, is-friend, similar, reliable, capable and negative labels like unreliable, incapable or untrustworthy. Some endorsements are static in that, once applied, they don't change over the course of the simulation (e.g. is-kin), while others are dynamic and may be revoked or replaced according to an agent's experiences (Werth et al. 2007). All agents use the same list of endorsements but differ in how they assess them.

To do so, agents rely on a so-called endorsement scheme which associates each label with a weight to express how much store an agent sets by this particular aspect of a person. Weights are modelled as integer numbers between *1* and *n* for positive labels and -1 and

[1] Endorsements were first devised by Paul Cohen (1985) as a device for resolving conflicts in rule-based expert systems. Scott Moss (1995) modified and extended their use within a model of learning by social agents. This latter version of endorsements has been adapted for the declarative model.

$-n$ for negative labels, respectively. This allows for computing an overall endorsement value for a person, applying the following formula:

$$E = \sum_{w_i > 0} b^{w_i} - \sum_{w_i < 0} b^{|w_i|}$$

where b is a number base and w_i the weight of the i^{th} label. Agents are assigned random endorsements schemes at creation, which differ not only in the weights they assign to the labels but also in the values used for n and b.

The overall endorsement value allows an endorsing agent to choose the preferred one(s) among a number of endorsees. To form the friendship network, for example, agents first determine which other agents are similar to themselves. This is based on abstract tags (to model character traits), which are assigned randomly to agents at creation. These tags are used to compute a similarity index (the number of similar tags), which in turn is used to generate "similar" and "most-similar" endorsements. Agents then compute the endorsement value for all known other agents and choose the ones with the highest values as their friends. Similar processes are applied to decide upon sexual partners and applicants for piece jobs.

There have been challenges in ensuring that modelling activities address the methodological challenges they set out to explore, while at the same time representing the case study material in a way that integrates multiple stresses and social networks and that can be validated by stakeholders. These challenges are being addressed via an iterative interaction between the field researchers and modellers and the field researchers and stakeholders. It is anticipated that this model of interaction between modellers and stakeholders enables a more representative capturing of stakeholder attributes in the constructed models.

Part of the initial development of model agents was based on ontologies developed between the Macaulay modelling team and the field researchers. Ontologies are defined by Gruber (1993) as "formal, explicit specifications of a shared conceptualisation": formal in that they are machine processable; and explicit in that all pertinent concepts and relationships between them are represented.

The methodology entailed assembling evidence from existing written work on the case study area which enabled subclasses, data types, object properties and supporting classes to be defined. The ontology was then tested for consistency using reasoning software. An iterative process refined the ontology and allowed new information to be brought in.

In the South African case study the ontology drew on research around smallholder farmers and climate information dissemination following the South African food security crisis of 2002-03. The ontology focused on describing various classes of farmers as discussed in the research, and in particular included definitions of qualitative but politically sensitive terms such as 'vulnerable farmers'. The ontology provided a useful tool to help field researchers qualify what they had meant in some of the research and provided the modellers with a comprehensive object model, defining detailed attributes of agents and other entities. On the other hand, it did not specify any actions of the agents, be they reactive (responses to a certain stimulus) or pro-active (pursuing goal-directed strategies). It was concluded that ontologies are a good tool to capture the static structure of a model domain but are not very well suited to model dynamic processes like agent's behaviours.

3.5 Validation data

There are profound problems with ensuring that data at the local level in South Africa are representative. The initial data and conceptual understandings used in the ontology were drawn from previous research, in what amounts to a TAPAS (Take a Previous Model and Add Something) approach to model design. Sekhukhune was deliberately chosen due to the activities of previous studies in the region. The FIVIMS (Food Insecurity and Vulnerability Information Mapping Systems)[2] project provided both qualitative and quantitative data and links were made with the RADAR (Rural AIDS and Development Action Research Programme)[3] project, which has data related to HIV/AIDS and social networks, so can be considered as a cross validation process. This approach was formalised through a meeting in Oxford in June 2006, which brought together the Manchester Metropolitan University modellers, the South African CAVES case study researchers, a RADAR staff representative (linked to University of Witwatersrand and London School of Hygiene and Tropical Medicine) and a FIVIMS researcher, with experience in food security policy.

Therefore, macro socioeconomic data were obtained largely through the municipal contacts in Sekhukhune and from nationally available statistics and research results, whereas micro-behavioural data were obtained during the field visits to the village and supplemented by other village-level research from RADAR that had been undertaken in the area. The outputs included both qualitative and quantitative information. The latter included detailed information about household structure, age distribution, income and migrants, which we were able to extract from several surveys conducted in the context of the IMAGE[4] study in eight villages in Sekhukhune. In terms of information about social networks, general data about the nature of networks (kinship and household networks, savings and funeral clubs) their evolution and social role were instrumental.

On the other hand, social networks datasets were also collected but proved quite difficult to use. While they were not comprehensive enough for comparison with model output on friendship networks in the validation process, they nevertheless helped to identify types of relationships (friends are relatives, neighbours, colleagues, known from school, church or savings clubs). Moreover, we discovered that people in the village do not distinguish between friends and acquaintances – so far one of the assumptions of the models.

---

[2] http://www.fivims.net/

[3] http://www.wits.ac.za/radar/Home.htm

[4] Intervention with Microfinance for AIDS & Gender Equity, see http://www.wits.ac.za/radar/IMAGE_study.htm

As noted above, the data requirements were framed to a large extent by the modelling and analysis of simulations, and these requirements were then addressed by the planning of data collection and validation exercises as detailed in the next section.

3.6 Development of stakeholder/data validation exercises and outcomes

The pilot-phase of fieldwork in Sekhukhune district was undertaken in February 2006 and highlighted some key issues that were integrated in the first round of modelling. The findings from the fieldwork were then presented to a range of stakeholders in April 2007 for validation purposes and for further development of the model. Both village and municipal stakeholders were approached for verification and clarification of data collected and to test assumptions made during the previous research phase. This helped to establish areas that required further secondary information and also guided the development of the agent-based model further.

Stakeholder feedback on these issues helped to establish which trends and issues identified in the fieldwork actually resonated with the communities' perceptions in a type of 'participatory validation' process over three days in the village. First a feedback meeting was held that was open to the whole village. Then focus groups were held with separate men's and women's groups. Some individual interviews were held and there were a number of visits to communal projects and village activities. A municipal meeting was also held to feed back results to the municipality and to allow the village representatives and municipal officials to have a discussion about the village results and the municipalities' role. This is a type of validation that can be referred to as 'generative reasoning'. The modelling process involves deductive reasoning, while, generative reasoning in ABM refers to an iterative process between inductive reasoning and deductive reasoning. There will be the opportunity to explore how this process may be integrated into future planning.

At the village level: The feedback meeting to the village was attended by about 100 people. Results from previous fieldwork in 2006 were presented using flip chart sheets. The main issues under 6 headings/themes were presented, namely: water, climate, food, health, jobs and communication; comparing what people in the village had said about the issue to how the municipality was thinking and talking about the issue (capturing the scale effect and also different priorities). One topic was presented at a time and then opened up for discussion and comment. The response was mainly positive in that people felt the key points experienced in the village had been well captured. People went on to place emphasis or expand on certain points and on numerous occasions disagreed with what we were reporting the municipality had said. The process provided a good way of verifying our findings and it highlighted the difference in views and priorities between the village and municipality and the tension that exists between them. People also commented on how pleased they were that we had come back to talk to them about what we had done and found and mentioned their frustration with government for not doing the same.

A visit to the community garden, illustrated that they are still struggling to get it off the ground as water is not accessible. A focus group meeting with both the home-based care group and the child care group was distressing, as they reported that there are 64 child-headed households in the Ga-Selala village.

At the municipality level: The meeting with Greater Tubatse Municipality was welcomed and our report was much appreciated. People hoped it could be used to focus attention from different departments and from the provincial level on the key issues highlighted in the report.

For part of the meeting, four village representatives joined the municipal meeting. The findings were presented using the same themes and points and then time was allowed for the village representatives and the municipality officials to offer their views on these issues. Certain issues were raised that allowed the village representatives to express their frustration with the municipality particularly linked to basic services and jobs. The municipality tried to explain their difficulties addressing certain challenges. It was a useful exercise in generating multi-stakeholder dialogue and one that can hopefully be built on.

The final fieldwork phase of the case study took place in October 2007. The objective was to present the results of the research to the same groups of people - the village communities and the municipal representatives - as visited previously, and it involved discussing model macro-results and model-derived storylines. The storylines were produced from output files from the declarative model by developing so-called 'traces' from social processes in the simulations. Traces were produced by selecting interesting or curious episodes in the output. Storyline were written up from the traces in order to present more human-understandable forms of the traces that could be read out/translated in the presence of the participants. These storylines could also be transformed into questions for further discussion and, in particular, validation.

Whereas the storylines micro validation was carried out with villagers, the macro-data validation exercises were carried out with municipal participants. In this case, simulations with the declarative model, which focused on investigating the impact of mining on the village communities, provided material for the validation activities. Time series macro outputs of the social simulation were used. Here, presenting the real plots might have been misread to suggest that the models produced some usable forecasting outcomes; therefore, schematic diagrams were used instead.

The CAVES modelling process is intended to contribute to policy-relevant debates. The above focus is pertinent to current policy and intervention activities in South Africa. It is hoped that the CAVES model output will provide lessons and insight into food insecurity dynamics that can be fed into policy networks such as "Renewal", SAVI (Southern Africa Vulnerability Initiative), the South African Government Social Cluster Food Security Group and FIVIMS-ZA.

4 Poland: Land reclamation in the Odra River

4.1 Objective of the case study

The Odra case study addresses the issues of land and water use and collective action with reference to land reclamation system (LRS) maintenance. In the Odra, land reclamation contains important environmental and socio-economic dimensions and provides insight into the role of social networks (collective action of neighbours along canals) in land-use (functioning of land reclamation affects the yields). In this case study these issues have been captured in an Agent Based Model of social maintenance of land reclamation system coupled with a biophysical model of hydrological processes and crop yields. Policy relevance of LRS issues arises from the significance they gain in the perceptions of stakeholders.

4.2 Brief description of the domain

The Odra land reclamation system consists of a network of drainage pipes, ditches and canals, sluice gates and fish ponds that drain land, particularly during the early spring thaw, in order to facilitate crop production. The better-maintained canals and sluice gates also serve as an irrigation resource. The primary purpose of the fish ponds is fish farming, however, under the national water law they are viewed as a component of the LRS and serve a key role in the drainage system.

Agriculture has been in decline in Poland since the socio-political reforms in the late-1980's. Lower profit margins in conjunction with high interest rates resulted in sustained under-investment in agricultural infrastructure. During the same period responsibility for LRS maintenance was transferred from government to individual land users. Although land users could have been forced, under the Water Law Act, to undertake the required cleaning and mowing of canals and ditches, in general the system has suffered neglect. Local officials charged with enforcing The Law have often overlooked this neglect out of "pity for the farmers" (because people are poor and life is hard) or simply in an endeavour to win favour and be re-elected. As a result fishpond owners have become almost exclusively responsible for LRS maintenance.

The case-study research indicates that economic, personal and institutional drivers are responsible for the current state of the LRS. Since decline in one part of a canal influences the rest of it detrimentally, system maintenance requires a collective effort. This interdependency is not symmetrical and varies largely in respect to the location along the canal. The direction of interaction depends on whether the LRS is serving a drainage or an irrigation function. In the case of drainage the negative impact of maintenance-neglect can spread upstream. If a downstream agent fails to maintain their section of the channel, the drainage function, and ability to cultivate crops, above this section will be disrupted. Equally, however, if an upstream land user harvests canal water using a sluice gate, they expose downstream farmers to water deficits and losses during dry periods. The collection of water in fish ponds has the same effect.

Prior to 1989, obligatory, top-down Water Partnerships (WPs) were responsible for LRS maintenance. Subsequent efforts to transform WPs into “bottom up” institutions have failed. Farmers neither want to pay for WPs not take private responsibility for LRS maintenance. At the same time farmers complain that the LRS is declining and perceive this fact as a real problem.

4.3 Stakeholder participation activities and emergent findings

Stakeholder participation allows the Odra case study to achieve the level of policy relevance required by the CAVES programmes. Workshops with local communities and stakeholders were used to inform the research and modelling tasks. In other words, the choice of the modelling subject resulted exactly from the workshops' findings with regard to the most important problems in the study area. The decline of the LRS was pointed out as the most severe land use related problem. Respective workshops were held in Zaborow and Kwiatkowice villages in order to elicit knowledge on local problems. Every resident of the village and the local authorities were invited and they were main participants. Workshops were facilitated by a local NGO, very experienced in work with these local communities. Without such cooperation it would have been much more difficult to perform any stakeholder participation activities. Attendants were asked to write down three main problems of the region. The results showed flood risk and LRS maintenance to be the chief concerns. In addition key experts in the fields of land use, water management and social welfare were consulted in a workshop held in Wroclaw in order to elicit knowledge on main regional development determinants and to discuss problems pointed out by local communities.

Workshops were augmented by fieldtrips, expert meetings and key informant interviews which were used to verify workshop data and deepen the research teams’ understanding of the domain. This understanding has been applied in the:

- design of both policy relevant and 'CAVES relevant' ABM modelling.
- drawing-up of a semi-structured interview questionnaire aimed at eliciting decision rules regarding both land use and land reclamation maintenance and understanding the role of social networks in the formation of these decision rules.
- selection of Rogow Legnicki village as a representative area for detailed research, GIS inventory and biophysical model development.

The farming household was used as the sampling unit of this research and was attributed as interview "respondent". Every household has been given a number (owner_code) in the GIS data base. Every interview is marked with corresponding owner_code. This enables the researchers to make interview results (and storylines) consistent with landownership/land use maps.

After the first phase of research in Rogow Legnicki the interview questionnaire was modified. Modified interviews were performed in two additional villages, Kawice and Kwiatkowice, both in the nearest surrounding of Rogow Legnicki. The third, and final wave of interviews was performed in the Peclaw area, where LRS is maintained properly and thus it might be a good reference for comparisons with Rogow L. area. Similarly to the Grampian case study, the iterative approach to field research and modelling was applied and for the same reasons.

4.4 Design of agents in case study model(s)

The workshops and fieldwork undertaken in the initial stages of this case-study identified important relationships and roles within the LRS. These findings formed the basis of agent construction and agent design. In this chapter we focus on agents design from the point of view of field research.

Agents in the Odra model represent households, not as a one-to-one mapping but as a generalized agent type.

So far there are two models being developed: a more stylised, coarse-grained model just inspired by field research findings and a more detailed, fine-grained model, more accurately grounded in research findings.

4.4.1 Coarse grain model (SoNARe)

The current version of the model considers two types of actors: the farmer, i.e. a general prototypical farmer type, and the water partnership initiator (WPI). Farmer agents are embedded in a dependency network according to the location of their land parcel along their respective channel and in an overall acquaintance network which is randomly superimposed on the dependency networks thus spanning all channels. The acquaintance network not only includes all farmer agents, but also a WPI agent, which is linked to all farmer agents.

In order to capture the main characteristics of the evidence obtained from the stakeholder interviews and to make a first feasible abstraction of the compiled decision rule set, the agent-based model focuses on explicitly contrasting social and economic influences on the decision making of farmers with regard to LRS maintenance. On the basis of these two factors farmer agents decide on their LRS maintenance strategy, changing it if their perceived economic success and their perceived social support fall below a certain level:

```
if(socialSupport + economicSuccess < 0) switch_LRS_strategy();
for(allAcquaintedFarmers) {
   if(otherLRSStrategy == myLRSStrategy) then
      supportOther(AMOUNT);
   else putPressureOnOther(AMOUNT);
}
```

Additionally a WPI agent exerts its social influence in favour of LRS maintenance once it perceives at least three farmers who have big losses.

4.4.2 Fine grain model

Agent types were designed based on elicited knowledge gained in the form of the storylines derived from interview transcripts and include the following types:

- Unwilling part-time farmer (UPTF),
- Willing part-time farmer (WPTF),
- Big farmer without fish pond (BFNP),
- Fish pond owner (FPO), WP initiator (WPI).

The research team also agreed that the Water Partnership was an institutional agent.

The inclusion of household agent characteristics and distinctions refer to land ownership issues, fish farming, activity related to Land Reclamation System maintenance, agricultural knowledge and social network integration. These criteria constitute agents' properties and their values are attributed to agents' types as shown in Table 1.

The set of agents and their properties were validated with one domain expert by one-by-one checking whether defined agent types and their features are probable/present in reality. It is a matter of further consideration of the modelling team which agents to include into the model.

Table 1: Property values of agent types in the model -- UPTF - Unwilling part-time farmer, BFNP - Big farmer without fish pond, FPO - Fish pond owner, WPI - WP initiator, WPTF - Willing part-time farmer; + must have this property value, - must not have this property value

Agent property	*Property values*	*UPTF*	*BFNP*	*FPO*	*WPI*	*WPTF*
Land ownership	Big farm		+			
	Small farm	+		-		+
	No farmland					
Fish farming	Yes			+		
	No	+	+			+
Activity related to LRS maintenance	Take leadership				+	
	Member of or willing to join WP		+	+	++	++
	Not member of WP but cleaning ditches	-	+	+		-
	Do nothing but willing		-	-		+
	Do nothing and unwilling	+	-	-		
Agricultural knowledge	High		+		+	
	Middle					+
	Low	+				
Social network integration	High				+	
	Low					

4.5 Validation data

The initial work undertaken on this case-study has assisted the researchers in identifying a region and stakeholders and informants that are reasonably representative, and which the researchers are able to place within a broader context. This represents an important first step in the model validation.

The representivity of the study area and population should be considered in relation to the research and modelling problem. The Odra case study deals with land use and sociological and biophysical aspects of land reclamation system (LRS). The Odra case study consists of 18 communes, of which only a part (the Rogow Legnicki area) of one (Prochowice commune) was researched in detail and used for model development. It is important to reflect on the extent to which the selected study area is representative of the broader area. "Key domain experts" were used to establish that the selected study area presents an accurate representation of the typical Polish situation and problems pertaining to LRS. In addition the analysis of indicators (Table 2) connected with farming and land use proves that the study area Rogow Legnicki is similar to the rest of the Odra.

Table 2: Comparison of land use parameters in study area of Rogow Legnicki and the broader Odra

Representativeness indicators	*Average value for Odra case study (rural areas)*	*Study and modelling area (Rogow Legnicki)*
Land use structure [%]		
farming grounds per total area	62,47	68,10
arable lands per total area	52,42	53,94
grasslands per total area	9,81	14,16
forests per total area	25,87	27,39
arable lands per farming grounds	83,55	79,20
grasslands per farming lands	16,06	20,79
Farms acreage [%]		
≤1 hectare	39,87	44,59
(1ha, 2ha)	17,96	9,46
[2ha, 5ha)	15,14	17,57
[5ha, 7ha)	5,26	6,76
[7ha, 10ha)	6,29	9,46
[10ha, 15ha)	5,44	5,41
[15ha, 20ha)	2,18	4,05

[20ha, 50ha)	3,27	2,7
[50ha, 100ha)	1,29	0
≥100ha	0,428	0
Farming structure		
Proportion of farming grounds in individual farms [%]3	70,14	70,55

Detailed boundaries of the area were established so as to meet the requirements of hydrological coherence. A reasonable sub-catchment was separated based on a hydrographic map of the study area.

The parameters used to make this comparison and establish the validity of the research site were sourced from a variety of data-sets, all them known to be credible.

- for land use and farming structure in the Odra case study generally: Regional Data Bank (RDB), Central Statistical Office;
- for land use and farming structure in study and modeling area (Rogow Legnicki): GIS data derived from 1:5000 land use and ownership map and associated abstract form land ownership registry;
- for farms acreage in Odra case study generally: Agricultural Census (AC) 2002, Central Statistical Office;
- for farms acreage in study and modeling area (Rogow Legnicki): GIS data base derived from 1:5,000 land use and ownership maps and associated abstract form land ownership registry.

Households were selected as the sampling unit. The small population in the study area allowed each land-owning household to be interviewed. Households were identified using land title registers (which includes names, addresses and parcels' numbers) obtained from local authorities. Where possible the interviews were specifically conducted with the person responsible for on-farm decisions. The credibility of the information gathered is to be verified by experts.

As regards availability of historical data, both climatic and economic data for the last 50 years are easily accessible. It is possible to analyse climatic time series data in order to choose specific periods influencing land use in certain ways. The same method can be applied to economic data including profitability of different types of agricultural production. The next step would be the spectral analysis of economic, social or agricultural time series data (like yields or land use structure), performed to extract such elements of real time series such as trends, periodicity and noise and compare them with respective elements of time series obtained as a model output.

In the case of hypothetical situations, time series are not available for scenarios or real system reaction and validation will rely on experts' intuition rather than their knowledge.

The scenarios of climate change might be created using global scenarios of climate change in order to modify weather time series from last 50 years.

Networks design

Using the data, two types of networks were constructed: family network and neighbourhood network. The information on this networks was formalized in the form of Excell spreadsheets that include following data:

i. OWNER_COD - an unique number representing given agent (household or institution/company) that owns the land.
ii. DIRECT_NEIGHBORS – codes of agents, who are direct neighbors of a given agent (in the sense of neighboring houses NOT fields); in case of Rogow Legnicki inhabitants we point direct neighbors codes known from addresses and the village map; in case of agents from other locations we give the location's name only; in some cases we don't know the address so it's impossible to indicate direct neighbours.
iii. FAMILY_TO – codes of agents related with a given agent.
iv. MANY_SOC_TIES – from interviews and/or observation we have information that enables us to indicate agents that probably have many social ties due to their position in the local community or/and type of profession; these agents are pointed out in this column with +++ and a comment.
v. AGENT_TYPE – private household or institution/company.
vi. WELLKEPT_HOUSES – on the basis of field observation we point out especially well kept houses with +++, which may indicate especially active and smart agents

In order to construct the neighbourhood network in the sense of neighbouring houses, we made a field trip to Rogow Legnicki and drew a map of the village with houses' location and numbers. Direct neighbours of every agent were identified according to our knowledge of household distribution in the village. Since we have a database with names and addresses, it was easy to convert house numbers into agents' codes.

A neighbourhood network could be also defined through neighbouring land parcels, which may be useful in our case study. Data on this kind may be easily derived from landownership map.

These two kinds of networks are not necessarily the same.

From field research and data from a document "Abstract form land ownership registry" provided by local authorities, we were able to identify some of the family ties within Rogow Legnicki village. This served as both a means of constructing the model validating the construction of the model where alternative data sources were used. Names of every landowner's parents can be found in this registry. In some cases this made it possible to deduce the exact nature of relationship (parents, sister etc.). In some cases this was not possible and only common family names and/or addresses indicate that some kind of family relationship exists. In few cases we know something about family relationships from field research, however not directly from interviews. This information comes from informal conversations and chats of our interviewers with local people, which seems to be more efficient method of eliciting this kind of information than forthright questions about these issues in the questionnaire.

4.6 Development of stakeholder/data validation exercises and outcomes

Three levels of stakeholder/data validation are possible in the Odra:

1) Validation of research results, which should be conducted following approved scientific standards aimed at establishing the "reproducibility" of the results. Another task will involve checking the interviewed farmers' credibility by comparing their statements with the point of view of experts – a process that has already begun. This is necessary so as not to be biased by the attitudes of one group of stakeholders, even if their decisions are basis for agents design.
2) Validation of rules, to check whether translation of knowledge from natural language to formal modelling language (to create an understanding of agents' real behaviour) was correct. Expert validation of modelling assumptions will include a set of questions based on a set of rules as shown in Table 3. Key domain experts that have been engaged in cooperation with the Polish CAVES team so far will be asked to validate rules. Farmers may be used in model validation provided it is possible to consult the same people that were approached in the initial fieldwork. Such stakeholder validation would have to deal with the problem of farmers misinterpreting the aim of the modelling exercise.

Table 3: Example of translation of rule to natural language question

Rule for BFNP	Questions for an expert
IF (haveGrassLand OR haveFallowLand) AND suffEconomicBalance THEN buildPond	Is it probable or reasonable to assume that a big farmer would build fishpond if he had some grassland or fallow land and enough money to do that? OR Does it really happen that big farmers usually build fishponds if they have grasslands and they can afford it?

3) Validation of modelling outputs should provide information on whether the reaction of the virtual system for different scenarios is consistent with:
   - either knowledge on the real reaction of the system being modelled for the scenarios that occurred in reality,
   - or at least intuition about possible reaction of the system to scenarios that never took place before.

In case model output is not detailed enough to compare it with real data series and/or data series on relevant level of aggregation are not available we must rely on expert's assessment of simulation outputs. The output of the biophysical part of the model has already been validated in that way and according to expert's opinion seems to be plausible. It is still a matter of discussion how many experts in validation process is enough.

Role-playing game as a tool for validation

The game was developed so as to meet the requirements of both a model validation exercise and a sociological experiment. The idea was to use this game to validate the model that is being developed by the Kassel team. The game is built on top of the SHAM biophysical model that we have developed for the Kassel model, and a simple economic model. To make the game more attractive, we have designed and created a set of graphics that represent various game concepts like economic condition, profit, or yield. The game is designed for a network play. Each player has a computer and connects to the game server via a local connection.

Each player of the game plays a role of a farmer, who owns a parcel. There are six such parcels in the game, each owned by a different farmer. The parcels lay on a piece of land of small and homogeneous slope, along a homogeneous channel, that runs through the centre of every parcel. The parcels' area and land use is homogeneous. The game works with a yearly time step and in each time step it calculates yields and a simple economic balance for every parcel. At the start of the game, each farmer has the same amount of money. During the game farmers spend their money on agricultural production and optionally on maintaining the ditch. They earn by selling agricultural products. A player can perform certain operations, like cleaning their part of the ditch, or filing a complaint against a neighbour, who does not clean their part of the ditch. A player against whom such a complaint is filed will have to pay a penalty in the next time step.

The gaming exercise was performed three times in three different villages with invited farmers. We engaged students of sociology to support farmers in technical staff and to make additional interviews concerning both gaming and real strategies of LRS maintenance. The results of the game are now being analyzed.

5 Conclusion

This report represents work in progress. The CAVES project has three case studies, Grampian, Sekhukhune, and Odra.

Comparing the three case studies, a number of conclusions can be drawn concerning participatory research techniques, model design and validation – points iii) to vi) describing the use of validation protocols. Some case study teams adopt validation protocols from the complementary CAVES conceptual paper deliverable 8. For example, the TAPAS approach was used in the Grampian and Sekhukhune case studies.

All cases combined existing datasets with primary data collection and utilised both qualitative and quantitative data. All case studies use participatory research techniques to collect and validate input and output data. Validation data were used more extensively as input to modelling rather than as data to compare with model outputs. This reflects a consensus (discussed at length in CAVES deliverable 8) that outputs should not be considered to be predictions of the likely occurrences of specific events. Outputs in terms of the statistical properties in time series, spatial distributions etc. could however be usefully compared. For example the Grampian study identified changing farm boundaries, farm size distribution and land price outputs. Sekhukhune and Odra cases relied on participatory approaches to validating outputs, in this case suitable datasets being unavailable or not comparable with model generated ones.

Another finding, consistent in all cases, is the difficulty of collecting social network data that would be useful to inform modelling. Network data were, however, used in the Odra study to construct neighbourhood and kinship networks (section 4.5). In the Sekhukhune model, different types of network were produced by the dynamic formation and dissolution of relationships. Validation concerned identifying the different types of relationships rather than analysing structure of networks.

Differences in the protocols and the validation techniques actually selected are also due to differences in the objectives and emphasis placed in the cases. These approaches have been considered within the case study teams with specific objectives in mind, but also as part of an ongoing dialogue across the case studies and project-wide. The variety of participation, design and validation methods reported in the previous sections that compose the validation protocols are a subset of those identified in the project meeting in Vienna (CAVES deliverable 8, section 5). The identified activities provided both a helpful checklist, to establish that different categories of validation had been considered, and overview of activities useful to focus on which ones should be the most important. The list was referred back to and reasons listed as to why activities were or were not pursued.

Although each case study team uses different techniques (such as expert validation and ontology) in implementing their specific validation approach, all members are working towards the aim of the policy relevant ABM. One reason this has been achievable – and a common research protocol across all case studies - is the iterative grounding of modelling in validation activities and fieldwork research. This allows the modelling work to become more relevant and focused on policy related issues.

6 References

Aamodt, A. & Plaza, E. (1994) Case-Based Reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications* 7 (1), 39-59.

Burton, R. (2004) Seeing through the 'good farmer's' eyes: Towards developing an understanding of the social, symbolic value of 'productivist' behaviour. *Sociologia Ruralis* 44 (2), 195-250.

Cohen, P. (1985): Heuristic Reasoning – An Artificial Intelligence Approach. Boston, Pitman Advanced Publishing Program.

Gruber, T.R. (1993) A translation approach to portable ontology specifications, *Knowledge Acquisition*, **5**(2):199-220, 1993.

Moss, S. (1995): Control metaphors in the modelling of decision-making behaviour. Computational Economics, v. 8, pp 283-301

Polhill, J. G. & Izquierdo, L. R. (2007) FEARLUS Model 1-0 User Guide. Technical report available from the authors.

Polhill, J. G., Parker, D. C. & Gotts, N. M. (2005) Introducing land markets to an agent-based model of land use change: A design. In Troitzsch, K. G. (ed.) *Representing Social Reality: Pre-Proceedings of the Third Conference of the European Social Simulation Association, Koblenz, September 5-9, 2005.* Koblenz: Verlag Dietmar Fölbach. pp. 150-157.

Polhill, J. G., Parker, D. C. & Gotts, N. M. (2008) Effects of land markets on competition between innovators and imitators in land use: Results from FEARLUS-ELMM. In Edmonds, B., Hernandez Iglesias C. & Troitzsch, K. G. (eds.) *Social Simulation: Technologies, Advances, and New Discoveries*. Chapter 7.

Simon, H. (1955) A behavioral model of rational choice. *The Quarterly Journal of Economics* **LXIX**, February, 99-118.

Werth, B.; Geller, A. and Meyer, R. (2007): He endorses me – He endorses me not – He endorses me... Contextualized reasoning in complex systems. Papers from the AAAI Fall Symposium: Emergent Agents and Socialities: Social and Organizational Aspects of Intelligence, AAAI Technical Report: FS-07-04, Washington, DC, November 8–11, 2007